日本語情動音声データベースの開発；脳機能イメージング研究への使用に適した情動音声刺激

著者	矢倉 晴子, 中川 誠司, 外池 光雄
雑誌名	大和大学研究紀要
巻	1
ページ	227-231
発行年	2015-03-16
URL	http://id.nii.ac.jp/1677/00000033/

Creative Commons : 表示 - 非営利 - 改変禁止
http://creativecommons.org/licenses/by-nc-nd/3.0/deed.ja

大和大学　研究紀要　第 1 巻　2015 年 3 月　pp.227 ～ 231

日本語情動音声データベースの開発
；脳機能イメージング研究への使用に適した情動音声刺激

Development of a Database of Japanese Emotional Speech Suitable for Brain Function imaging-study

矢 倉 晴 子* 中 川 誠 司** 外 池 光 雄***
YAGURA Haruko　NAKAGAWA Seiji　TONOIKE Mitsuo

要　旨

最近，非侵襲的脳活動計測が発展し人間の「こころ」に関する研究が盛んになった。特に，相手の声から気持ちを推測するときの脳活動を検証するための情緒的プロソディ理解に関する脳活動が世界的に注目を浴びるようになったが，国内では研究が多くない。その理由として，脳科学に適した日本語情動音声データベースの不足がある。われわれは以下の点を工夫して音声データベースを作成した。

1）音声の感情的な印象にばらつきがあると実験結果に大きな影響を与えるので，あらかじめ音声の印象評価値実験を行い，聞き手がうけた感情的な印象の強度と種類（－3：最も否定的，0：どちらともいえない，＋3：最も肯定的）を明記したリストを各音声にリストとして添付し，実験者が自分の実験に必要な音声を選べるようにした。

2）音声の意味そのものが感情的であると，音声理解に加えて言語理解に関連する脳活動が同時に活動してしまうので，言語的な感情的意味合いが無感情である「日本語の苗字」で音声の意味内容を統制した。

人間が相手の声をきいて感情的であると印象を受ける時に，一番の手がかりになると報告されている音声の基本周波数と印表評価との相関を解析した結果，両者の間に中等度の相関が認められた（$r_s = -0.44$，$n = 145$，$P<0.001$）。今後，開発したデータベースを一般に公開し，国内で数少ない情緒的プロソディ理解に関する脳活動計測研究に貢献することを期待する。

keywords：情緒的プロソディ理解（Emotional Prosody Recogntion）・情動音声（Emotional Speech）・印象評価（Image Evaluation）

1．背景

1.1『ことば・こころ』のやりとりによって日々の会話がより豊かなものに

毎日取り交わす「おはよう」という挨拶でも，話者の声が低くゆっくりであれば「悲しく」聞こえ，明るく速ければ「楽しく」聞こえたりする。このように，『怒り』『悲しみ』など感情を表わす声のトーンあるいはイントネーションを総称して『情緒的プロソディ』と呼び声の速さ，強さ，高さで変化すると定義されている［1］。私たちは，この『情緒的プロソディ』を理解することによって，ことばの文字通りの意味に加え，状況によって多様に変化する相手の様々な「こころ」の動きを念頭において会話を行い，人間関係を円滑かつ個性あるものにすることができる。

それでは，日々の音声言語コミュニケーションで，情緒的プロソディ理解の能力を人間が失ったらどうなるのだろうか。情緒的プロソディ理解に関する機能は，右大脳基底核に局在するという報告が多く，その部位を脳卒中や交通事故などで損傷された患者さんは，目に見える体の傷は治癒したとしても，コミュニケーション上支障を生じ，「反応が悪い，何を考えているかわからない」などと誤解され，日常生活上で良好な人間関係を築く上で困難を生じると報告されている。このような患者さんたちは，病院を退院して，日常生活への支障を生じても，何のケアもされない場合が多い。この症状は，右半球損傷後のコミュニケーション症候群［1］［2］［3］などともよばれ，臨床や研究上のさらなる発展が必要である。そして，この機能がいかにわれわれの生活にとって重要であるかを考えさせられる。

1.2　情緒的プロソディ認知に関する脳科学的研究の世界的な増加

このように，我々の日々のコミュニケーションで欠かせない能力である『情緒的プロソディ』であるが，その脳内メカニズムの解明について近年 10 年の間に機能的核磁気共鳴画像法（functional magnetic resonance imaging: fMRI）などの非襲侵的脳機能計測技術の発展により，世界中で急速に論文数が増加した。『こころ』

* 大和大学医療保健学科　**（独）産業技術総合研究所　*** 藍野大学医療保健学部
平成 26 年 12 月 19 日受理

は個人差が多く，統一的なデータを得るための計測も難しい反面，『目に見えない』『こころ』を『可視化』するという，人類にとって非常に魅力的な分野であり，気分障害などの精神疾患関連の医療や，感性工学などの分野での応用の期待も大きく，今後一層発展していくと考える。

しかし，国内では『情緒的プロソディ』に関連する脳科学的研究が皆無である。その大きな理由の一つとして，脳活動計測に適した日本語の音声データベースがないことが考えられる。そのため，実験者が自作の情動音声刺激での実験を余儀なくされることになるが，音声を録音するための防音室などの収録環境や，話者の検討，音圧や基本周波数などの音響パラメータの統制に関する知識や技術，音声に対する聞き手の感じ方の個人差を統一するための主観評価実験にかかる手間など，さまざまな制約が発生する。

1.3 脳科学に適した日本語情動音声データベースの必要条件

情動音声データベースとは，できうる限り感情表現豊かに作られた音声である。しかし，Ericson,D. らの感情表現の種類と分類を扱った総説によれば，音声データが発話された状況の違いを考慮する必要があることを強調している。例えば，俳優などにより表現された典型的な感情表現を伴うのか，逆に，心から表現される自発的感情表現を扱うのか，などである。そして，研究の目的により，音声の何を研究すればよいのかが異なるため，研究の目的と発話音声の種類について，以下の３つに分類している［4］。

1）自発的発話の音響特徴の分析や理解であるならば，危険な状況のパイロットの会話や，面白いテレビを見た時の笑い声など，特定の感情を表出することが予想される状況により発話される自発的感情表現による音声。
2）研究の目的が，日常会話が持つ表現の豊かさの理解や，合成音への情報の付加ならば，自然な感情表現がついた自発的感情表現による音声。
3）アニメ，ナビゲーションなどの商用システムへの合成音声の提供ならば，感情を効果的に演出できるプロの俳優によって演じられた表現豊かな典型的な感情表現による音声。

脳科学実験に必要なのは，１‐３のどのタイプであるか検討するためには，脳の反応の特殊性について考える必要がある。脳波実験で代表的な事象関連電位（Event-related potencials: ERP）を例にとって考えると，ERP とは，大脳皮質における数百万のニューロンの活動によって生じる電位を頭皮上の複数の電極によって計測する脳波（Electoroencephalogram: EEG）の一種である。記憶や言語などのさまざまな認知処理に対応して，特定の潜時（ある事象から出現するまでの時間）に特徴的な波形成分（後述）が出現するため認知過程の測定に幅広く用いられる［5］［6］。しかし，これらの脳波は，入力される音の物理的特性によって，波形が大きな影響を受けるため，計測条件では，ノイズ処理など，厳密に環境を統制する必要がある。また，設定した条件間で，反応時間等の行動データをしっかりと得られない刺激は，明瞭な脳活動信号も十分に得られない。

そこで，第１に考慮する点として，刺激の分かりやすさである。被験者がその音声を聴いて，どんな感情をこめて話しているのかを容易に理解できないと，脳の反応が十分に生じない。言い換えると聴覚印象が不鮮明であり，被験者によって印象に個人差がある刺激は実験に不適切である。２）は，話者の主観に依存するある特定の日常場面で発話された音声を集めたものであるため，話者と聞き手の印象が必ずしも一致することはない。また，１）は特定の場面に依存するため，収録環境を統制することが難しく，音質にばらつきが生じる可能性があり，ノイズが脳の反応に影響を及ぼす可能性がある。そこで，本研究では，聞き手がわかるように刺激の聴覚印象を統制しやすく，収録環境も防音室など統制しやすい３）を採用した。多くの感情的プロソディ認知の脳科学的研究は，話者にある特定の感情をこめて発話させた音声を，実験音声刺激として使用している。その一例として，矢倉らは，声優などのプロの話者に，指定した感情を意図的にこめて発話してもらった音声を脳科学の実験に使用した結果，信頼性ある結果を得られている［7］［8］［9］。

第２に考慮する点として，事象関連電位計測をはじめ，多くの脳機能イメージング研究で加算平均という手法が用いられるということである。同系列の刺激による脳活動信号を 50 回や 100 回繰り返し，それらの信号を平均化することにより，精度の高い波形を得ることがいえる。この手法は，ミスマッチ課題，オドボール課題など，脳波にもっともよく用いられる実験パラダイムで一般的な計測方法となっている［6］。しかし，同じ刺激を数百回も繰り返し聞くことにより，実験時間が長くなり，被験者の疲労と脳の慣れによる信号の減衰などの問題が生じる。そのため，音声刺激はなるべく持続時間が短いものが望ましいと考える。そこで，日本語苗字で最も種類が多くかつ文字数の少ない３モーラで終わる語を選定した。矢倉らの実験で使用した音声刺激は，平均の持続時間が 0.74s（表１）で，1trial ３s 以内になり，被験者の反応時間等も含めて，50 回加算平均すると１条件 15 分程度で終了した［7］［8］［9］。これは，被験者が集中しうる妥当な時間であったと考える。また，それぞれの音声には，プロソディの３要素であるといわれ

る，基本周波数･音圧･持続時間の音響パラメータのデータリストを付与し，実験者が実験条件に合った音響パラメータの音声を選択できるように配慮した。

また，第３に考慮する点として，情動音声がもつ言語音としての意味の特性である。多く情緒的プロソディ認知の脳科学的実験では，刺激の感情的な意味に関しても厳密に統制している。たとえば語の意味のみに感情的意味合いを含ませ音声の感情的印象は無感情に設定した条件，意味を中立にして音声の聴覚印象のみを感情的に設定した条件，音声と意味の両方とも感情的な意味合いを含ませた条件，これらの脳活動信号データを各条件間で比較し，意味と音声の一致不一致の時の脳活動を計測している。しかし，多くの日本語で公開されている情動音声データベースは，自然言語処理の応用として，音声認識システムの開発などに使用されるため，さまざまな言語的意味をもった品詞が混在する言語内容で構成されている［10］［11］。しかし，脳機能イメージング研究で，理解する対象の単語の品詞の種類（カテゴリー）によっても脳が反応する部位が違うという報告も多く［12］［13］［14］，語のも厳密に統制する必要がある。そこで，本刺激では，刺激の感情的意味が中立である名前に音声すべての品詞を統一し，苗字＋さん（例；/araisan/）145 個を使用した。名前はどんな種類の感情をこめても自然に聞こえ［15］，日常会話でも身近な発話であり，被験者が理解しやすい，などの利点が挙げられる。また，５モーラの苗字で品詞とモーラを統制していることにより，持続時間などの音響パラメータと，音声の感情的印象の相関関係がより明確に抽出されうるという利点もある。

我々は，以下３つの点を考慮して，脳科学実験に適した，情動音声データベースを作成することを計画した。

１：話者をプロの俳優に依頼
２：持続時間を短くする，モーラ数を統制，音響パラメータのリストを添付
３：語のもつ感情的意味を中立，品詞と内容を名詞の日本語苗字に統制

２．情動音声データベースの作成

2.1　情動音声の収録

音声刺激は，すべて母音から始まる３モーラの苗字に，さらに２モーラからなる敬称「さん」の計５モーラからなる 145 種類の名前を使用した（例 /a･ra･i･sa･n/）。名前を使用した理由は前述したとおりである。音声刺激は，熟練した女性話者（28 歳）により，「肯定的；positive」「否定的；negative」「中立；neutral」の３つの感情的なトーンを込めて発話された音声 145 個（negtive;48 positive;50 neutral;47）を作成した。話者には，肯定的は喜び，否定的は悲しみ，中立は朗読，の３つの情動を意図して発話してもらうように指示した。音声の編集には，Cool Edit Pro, version 2.0 を使用した。音声ファイルは，44000Hz/16bit/Mono，Wav 形式でフォーマットを行った。音声刺激の収録は 2004 年～2006 年にかけて，産業技術総合研究所関西センター内の無響室にて行った。

2.2　主観評価値の収集と音響パラメータ値の分析

同じ音声を聞くにしても，感じ方に個人差があるのはどうしても避けられない。ある程度被験者全員が統一した印象をもつ刺激を脳活動計測に使用することにより，妥当な反応結果を得ることができる。また，脳は聞いてはっきりとした印象の刺激を入力しないと，明瞭な脳活動信号が得られない。そのため，感情的印象が強い刺激を実験用に選択する必要がある。これらのことを踏まえ，各情動音声には，情動印象の強度と種類を示す印象ラベルを付加することが必要になってくる。

さらに，音声の情動的印象と大きな関連を持つといわれている基本周波数や持続時間などの音響パラメータ値の情報も各音声に付加することにより，実験者が実験に必要な物理パラメータを持つ刺激を選択することができる。本刺激セットでは，各音声に主観評価値と基本周波数（Hz），持続時間（s），平均音圧（dB）の３つの値を付加した。音響パラメータの解析には，Wave surfer; ver.1.6.3, Matlab（ver.2006a）を使用した。

聞き手による音声から喚起される感情的な印象を調査するために，脳血管障害や耳鼻咽喉科疾患の既往がない健常者 30 名を対象に 145 種類の音声刺激聴取による印象評価実験を行った（女性 13 名，男性 10 名，年齢平均（歳）24.0，範囲 20-34）。被験者は全員，某国立大学の学部学生，大学院生であった。

	音響パラメーター		
	平均基本周波数	平均音圧	持続時間
	（Ｈｚ）	（ｄＢ）	（ｓ）
平均	311.0	47.3	0.74
ｓｄ	46.6	2.37	0.11

表１　音声の音響パラメーターの平均値（ｎ =145）

	平均基本周波数		平均音圧		持続時間
スピアマンの相関係数（両側）	0.44	＊＊＊	-0.54	＊＊＊	-0.02

表２　主観評価値と各音響パラメータとの相関

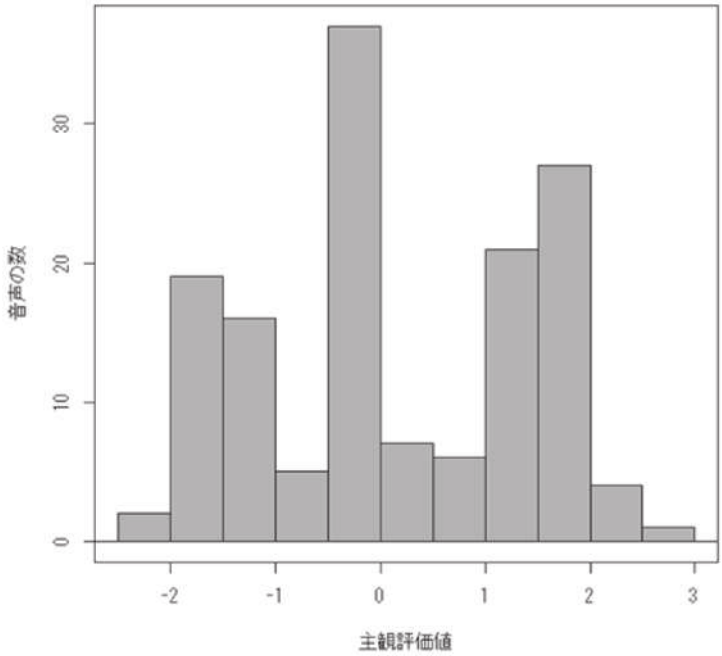

図１　音声の主観評価値の分布（左：ヒストグラム，右：度数分布表）

度数は，0（－３:最も否定的），2（－２:否定的），－１（35:どちらかというと否定的），0（どちらともいえない：42），１（どちらかというと肯定的：13），２（肯定的：48），５（３：最も肯定的）

実験者は被験者に以下の教示を行った。データの主な収集方法は，音声を CD-LOM で配布し，各被験者がヘッドフォンを使用して聴取するように指示した。なお，配布の際に，以下の教示について説明した。

『この声は，どのような印象で聞こえますか？用紙の記入欄に，もっとも否定的；－３，否定的；- 2, どちらかというと否定的；－１，どちらでもない；０，どちらかというと肯定的；１，肯定的；２，もっとも肯定的；+３のどれかの欄に，聞こえたとおりにしるしをつけてください，音声は２回繰り返します。２回で感じたとおりに記入してください。』主観評価の資料の収集は，2004 年 11 月～ 2005 年 1 月にかけて行った。

３．感情をこめて発話した呼名による日本語音声の特徴

１．印象評価値の分布

145 個の音声の主観評価値と音響パラメータ値の平均値を表 1 に，主観評価値のヒスト分布を図 1 に示す。主観評価値；-1（どちらかというと否定的）に印象が集中していることから，悲しみの印象は少なかったことが考えられる。反面，2（肯定的）に刺激のほぼ 1/3 が集中していることから，喜びの印象が強かったと考えられる。

２．印象評価値と音響パラメータの関係

音声の主観評価値と音響パラメータの関係を検証するために，相関係数の分析をした。その結果，主観評価値と基本周波数・平均音圧との間に中程度の相関が認められた(表２, スピアマン相関係数, 有意確率, 両側)。なお，相関係数の解析には統計ソフト IBM spss ver.19 を使用した。本刺激の特徴として，基本周波数と感情印象で正の相関が認められたことから，声が高くなると，印象が肯定的になることが認められた。反対に，音圧と印象に負の相関が認められ，音圧が高くなると，否定的な印象が増すことが考えられる。感情的な音声の聴覚印象を決定する音響特徴として，基本周波数，速さ，大きさ，声

質などがあり，特に重要であるのが，声質である［16］［17］［18］。声質の定義は，聴者が二つの音が同じラウドネス，同じ，ピッチであっても，異なると感じられる「音の質」であり基本周波数との関連が大きいと報告されている。本実験でも［16］［17］［18］，基本周波数と印象評価との関連性が認められたことから，感情音声データとして使用する妥当性を示唆する結果となった。

4. まとめ

脳機能計測での使用を想定した日本語の情動音声データベースを作成した。今後，これらを一般に公開し，国内の『情緒的プロソディ理解』に関する『ことばとこころ』の脳機能イメージング研究の普及に貢献することを期待する。

5. 参考文献

［1］ E. D. Ross, “The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere.” 1981.

［2］ 竹内愛子，高橋正，宮森孝史，“右半球損傷者のコミュニケーション能力，” 音声言語医学，vol. 30,pp. 178 － 187, 1989.

［3］ Mark D pell, “Judging emotion and attitudes from prosody following brain damage.” Progress in brain research, pp. 156;303—17, 2006.

［4］ D. Erickson, “Expressive speech: Production, perception and application to speech synthesis,” *Acoust. Sci. Technol*., vol. 26, pp. 317—325, 2005.

［5］ E. Potentials, “ERP（事象関連電位）データを読み解くための基礎知識 菅井 康祐 近畿大学，” pp.75—82, 2012.

［6］ 入戸野 宏，心理学のための事象関連電位ガイドブック. 北大路書房, 2005.

［7］ Y. Haruko, S. Nakagawa, Y. Kobayashi, S. Oginoa, and M. Tonoike, “Cortical activities associated with emotional prosody processing.” International Congress Series,Elsevier, p. 1278; 31—34, 2006.

［8］ C. Neurophysiology, H. Science, B. Engineering, A. I. Science, C. Author, H. Yagura, and A. H. Science, “MEG Measurement of Event-Related Brain Activity Evoked by Emotional Prosody Recognition,” vol. 89, pp. 1—5, 2004.

［9］ H. Yagura, M. Tonoike, M. Yamaguchi, S. Nakagawa, K. Sutani, and S. Ogino, “MEG measurement of event-related brain activity evoked by emotional prosody recognition.,” *Neurol. Clin. Neurophysiol*., vol. November, no.30, p. 89, Jan. 2004.

［10］ J. S. T. Crest, “The Recording of Emotional Speech-JST / CREST database research - Nick Campbell.”

［11］ Y. Arimoto, S. Ohno, and H. Iida, “Assessment of spontaneous emotional speech database toward emotion recognition: Intensity and similarity of perceived emotion from spontaneously expressed emotional speech,” *Acoust. Sci. Technol*., vol. 32, no. 1, pp. 26—29,2011.

［12］ E. Privman, Y. Nir, U. Kramer, S. Kipervasser, F. Andelman, M. Y. Neufeld, R. Mukamel, Y. Yeshurun, I. Fried, and R. Malach, “Enhanced Category Tuning Revealed by Intracranial Electroencephalograms in High-Order Human Visual Areas,” vol. 27, no. 23, pp. 6234—6242, 2007.

［13］ Z. Y. Q. Qing-Lin, “Specificity of Category-Related Areas in Ventral Visual Cortex.pdf,” *Adv. Psychol. Sci*., vol. 19, pp. 42—49, 2011.

［14］ T. J. Grabowski, H. Damasio, and A. R. Damasio, “Premotor and prefrontal correlates of category-related lexical retrieval.,” *Neuroimage*, vol. 7, no. 3, pp. 232—243, 1998.

［15］ 曽我部 優子，筧 一彦，河原 英紀，“感性情報に曖昧さがある場合の音声の心理的評価とその物理特性，” 電子情報通信学会技術研究報告，vol. SP, 音声 102, no.（749），pp. 37—42, 2003.

［16］ T. Brosch, D. Grandjean, D. Sander, and K. R. Scherer, “Behold the voice of wrath: Cross-modal modulation of visual attention by anger prosody,” *Cognition,* vol. 106, no. 3, pp. 1497—1503, Mar.2008.

［17］ K. R. Scherer, “Affect bursts.,” in *Emotions: Essays on emotion theory*., 1994, pp. 161—193.

［18］ J.-A. Bachorowski and M. J. Owren, “Sounds of Emotion,” *Ann. N. Y. Acad. Sci*., vol. 1000, no. 1, pp. 244—265, Jan. 2006.